# Identifying Possible Risk Factors for Dementia

Georgia Ettles, Mayez Haris, and Niharika Peddinenikalva
**University of Edinburgh**

30th November 2022

## Contents

# 1 Executive Summary

Dementia is a syndrome associated with the decline of one's general cognitive ability. It most commonly presents itself in the form of Alzheimer's disease, vascular dementia, and Lewy body dementia [9]. There are several possible risk factors for the onset of dementia with age, exercise, and depression being some of the largest [5]. The aim of this report was to explore the possible risk factors of dementia by constructing a statistical model and then use validation methods to check the accuracy of this model.

Participants' cognitive ability was measured using a composite cognitive score which was calculated by averaging the numeracy, recall and time orientation skills. This was presented on a scale from 0 to 29, with 0 and 29 representing low and high cognitive ability respectively.

It was observed that an increase in age by one year is associated with a decline in cognitive score by a factor of 0.14. Generally, men had a lower cognitive score than women, but as age increased, cognitive scores seemed to deteriorate at a similar rate for both sexes (fig. 1). Severe depression was also associated with a decline in cognitive score (fig. 2a), with those who were "Not Depressed" having a median cognitive score of 17.50 while those who were "Very Depressed" had a median cognitive score of 13.75. Another finding showed that participants who frequently partook in vigorous activities tended to have a higher cognitive score than those who didn't (fig. 2b), with rarely partaking in vigorous activities being associated with a decrease in cognitive score by a factor of 0.39.

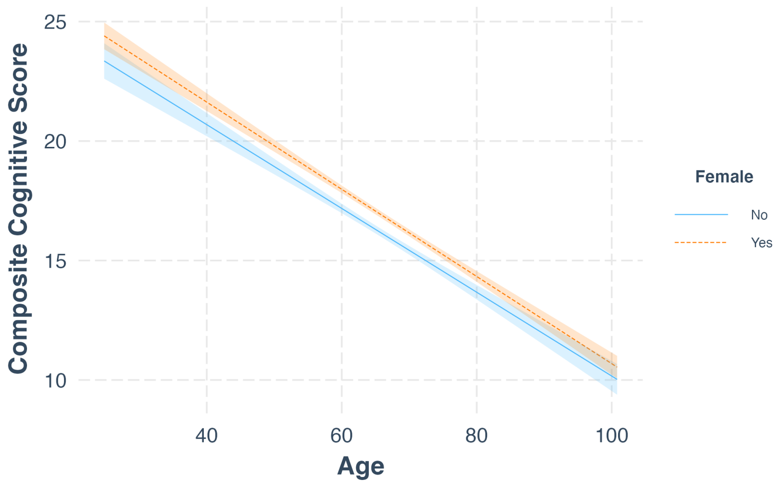
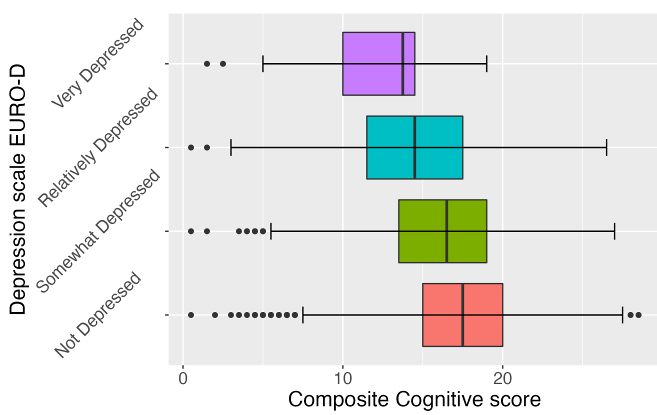
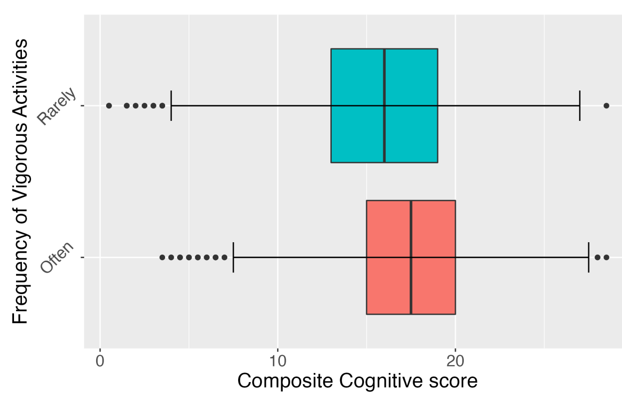



Figure 1: Visual representation of how cognitive score varies with age for both males and females.




(a) Depression Scale EURO-D




(b) Frequency of Vigorous Activities

Figure 2: Visual representation of how cognitive score varied for different levels of depression, frequency of vigorous activities in daily life and the number of chronic diseases one has.

## 2 Introduction

Dementia is a syndrome associated with a continuous decline in one's general cognitive ability severe enough for it to interfere with daily activities. This can present itself in the form of a few different diseases, with the most prominent being Alzheimer's disease, and some others including vascular dementia and Lewy body dementia [9]. The onset of dementia is thought to be primarily associated with an increase in age, [5], along with exercise and depression being possible risk factors when considering the likelihood of someone developing dementia [4]. Despite evidence of association with these risk factors, there is no proven cause for types of dementia like Alzheimer's or vascular dementia. The aim of this report is to use observational data obtained through a survey about various aspects of the participants' lives, and then explore this data to find possible risk factors for the onset of dementia.

# 3 Methods

## 3.1 The Data

The data used for this report was the easySHARE dataset published by SHARE [3]. The Release Guide to easySHARE (PDF) documents the data format and all variables included. The easySHARE release 8.0.0 data used for this report is based on SHARE Waves 4 [1] and 6 [2]. For more targeted analysis, the data for one country from one wave of interviews was chosen. A significant proportion of the data from the full dataset was found to be missing, thus to justify which combination of country and wave were chosen, the number of missing values and how many total observations there were for each country in each wave were checked. It was found that Estonia had a large number of observations in wave 4 after excluding each observation with a missing value. After choosing this dataset, all observations that had missing values were removed, leaving 5231 observations. Finally, 200 randomly selected observations were set aside which were used later to assess the constructed model. Thus, the "training" dataset used for the modelling was the remaining 5031 observations from wave 4 in Estonia.

There was no variable in the original data measuring overall cognitive ability. However, the data included various cognitive function indices, which were used to create a composite cognitive score, providing a more holistic representation of the participants' cognitive ability on a scale from 0 to 29, with 0 and 29 representing low and high cognitive ability respectively. This was an appropriate response variable for the constructed model. The average cognitive score of the participants from the selected data was 17.

Using information available online from medical organisations like the CDC [5] and the NHS [8], specific variables from the data were selected which could be possible risk factors for explaining the onset of dementia. In agreement with the health organisations mentioned and Livingston et al. [7], the biggest possible risk factors included age, gender, level of physical activity, depression, and more. Other possible risk factors which were thought to be important included alcohol consumption, smoking habits, mathematical skills at age 10, etc. However, due to large amounts of missing data, these possible risk factors were not considered in the statistical model.

Interaction terms such as age with the number of doctor visits was used to investigate how cognitive score varies with the number of doctor visits differently with age. The interaction terms are given in table 3. Also, variables such as BMI, number of children, marital status and interaction terms such as education with EURO-D depression scale were initially considered for the model. However, while ensuring that the linear model assumptions held, these terms were discarded due to their insignificance according to the AIC values during model selection with the function `step()` in `R`.

Since some of the variables in the data were named fairly cryptically, they were renamed for ease of understanding. For example, the variable `isced97_r` became `edu`. The rest of the changes made are shown in table 1.

| **Variable Name in easySHARE** | **New Variable Name** |
|---|---|
| dn004_mod | native |
| iv009_mod | loc |
| isced1997_r | edu |
| sp002_mod | outside_help |
| casp | life_quality |
| hc002_mod | doc_vis |
| hc029_ | nursing |
| br015_ | vig_act |
| ep005_ | job_situ |
| country_mod | country_name |

Table 1: The variables which were renamed for simplicity.

### 3.2 The Model

The final model for the composite cognitive score was a linear model with response variable $\text{cogscore}_i = Y_i$ for person $i = 1, \ldots, n$ given by

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i,$$

where $\epsilon_i \sim N(0, \sigma^2)$. $\mathbf{x}_i$ is a row of the model matrix for observation $i$ containing a continuous variable for age, dummy variables for factor variables $x_2$ to $x_{15}$, and variables for interaction terms $x_{16}$ to $x_{18}$ which are defined in table 2 and table 3 respectively.

| $x_j$ | **Variable name** | **Description of variable** | **Factor Levels** |
|---|---|---|---|
| $x_1$ | age | Age of participant | range of positive values |
| $x_2$ | female | Male or female | Male=0; Female =1 |
| $x_3$ | native | Born in country of interview | No=0; Yes=1 |
| $x_4$ | loc | Area of interview location | City_sub; Town_rural |
| $x_5$ | edu | Level of Education completed | Secondary and below; Post secondary; other = Still in school/other |
| $x_6$ | outside_help | Received help from outside the household | No=0; Yes=1 |
| $x_7$ | chronic_mod | Number of chronic diseases | 0; 1-3; 4-7; 8-10 chronic diseases |
| $x_8$ | life_quality | The CASP-12 score measures quality of life and is based on four subscales on control, autonomy, pleasure and self-realization. | Low; Moderate; High |
| $x_9$ | eurod | Depression scale of EURO-D | Not depressed, Somewhat depressed, Relatively depressed, Very depressed |
| $x_{10}$ | doc_vis | The number of doctor visits a participant has per year | 0-10; more than 10 doctor visits |
| $x_{11}$ | nursing | Is the participant in a nursing home | No=0, Yes=1 |
| $x_{12}$ | iadlza | Instrumental Activities of Daily Living Indices | Low Difficulty, High Difficulty |
| $x_{13}$ | maxgrip | Maximum of grip strength measure | Low; High |
| $x_{14}$ | vig_act | Frequency of Vigorous activities | Often; Rarely |
| $x_{15}$ | job_situ | Employment status | Retired; Employed: Unemployed; Sick Disabled; Homemaker; Other |

Table 2: List of explanatory variables.

| $x_j$ | **Variable name** |
|---|---|
| $x_{16}$ | loc:outside_help |
| $x_{17}$ | age:doc_vis |
| $x_{18}$ | nursing:iadlza |

Table 3: List of interaction terms within the model.

The matrix-vector form of the model is given by

$$E(Y) = X\boldsymbol{\beta},$$

where

$$X = \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_n^T \end{pmatrix}$$

is the model matrix,

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

is the parameter vector, and $\mathbf{x}_i^T = \begin{pmatrix} 1 & x_{i,1} & x_{i,2} & \dots & x_{i,18} \end{pmatrix}$. For example, $x_{i,1}$ corresponds to the age of observation $i$ and $x_{i,18}$ corresponds to the product of whether the participant is in a nursing home and the value of their instrumental activities of daily living index for observation $i$. Finally, $\boldsymbol{\beta}$ contains parameters corresponding to the model matrix $X$. The first level of each factor and the respective interactions was set to zero due to identifiability. The model was confirmed to be linear with almost constant variance by checking the Residuals vs. Fitted values plot. The remaining assumptions were also satisfied when checking the Normal Q-Q plot, the Scale-Location plot, and the Residuals vs Leverage plot (fig. 6 in appendix C).

### 3.3 Model Assessment

The predictive performance of the defined model was assessed by computing the root mean squared errors (RMSE) scores when the model was applied to different datasets. The RMSE scores check whether the model was accurate at predicting cognitive score for the training dataset along with other datasets. The model was applied to datasets for which the observations already existed so that we could investigate the variation of predicted values of cognitive score with the actual observational values. The datasets considered included

- Test 1: The test data was the same as the training data used to build the model, consisting of 5031 observations from Wave 4 in Estonia.
- Test 2: The test data included the 200 randomly selected observations which were initially set aside from Wave 4 in Estonia for the purpose of model assessment.
- Test 3: The test data included data from Austria (different country) in wave 4 (same wave) excluding any observations that had missing values.
- Test 4: The test data included data from Estonia (same country) in wave 6 (different wave) excluding any observations that had missing values.

The RMSE scores were calculated using the four test datasets listed above. Each test dataset had observed values of cognitive score for each observation, given by $y_i^0$. The model then predicted values of cognitive score for each observation in the test data, given by $\hat{y}_i^0$. The prediction errors (also known as mean squared errors) were computed as

$$PE_v = \frac{1}{n} \sum_{i=1}^{n} (y_i^0 - \hat{y}_i^0)^2,$$

where $n$ is the number of observations in the test data. Thus, the RMSE was calculated as RMSE $= \sqrt{PE_v}$.

To analyse the importance of variables in the model, The RMSE scores from the model were compared to those of a minimal model where cognitive score depended on age as the only explanatory variable. This minimal linear model is given by

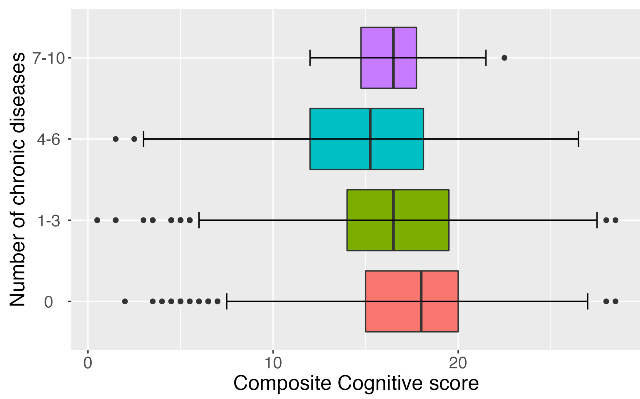
$$Y_i = \alpha + z_i \beta + \epsilon_i$$

where $Y_i$ and $z_i$ are the cognitive score and age respectively, and $\alpha, \beta$ are the regression coefficients for the minimal model. The RMSE scores for this minimal model were calculated for the four test datasets and compared with the main model.

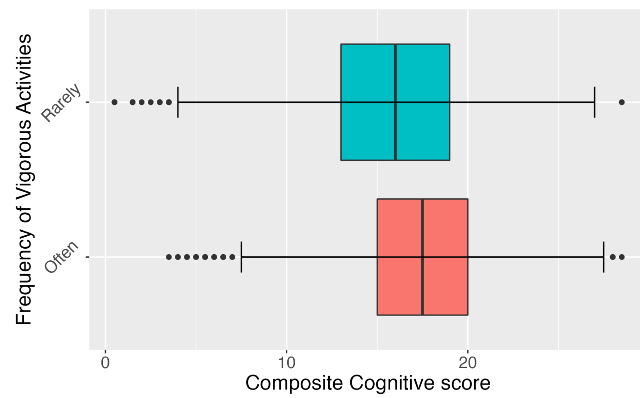
# 4 Results

The training data for Estonia in wave 4 had 5031 observations after all observations with missing data were deleted. There were a total of 1986 men and 3045 women. The average age of the participants was 66, with the youngest participant being roughly 25 years old and 20.69% being over the age of 75. Only 37.15% of participants had gained a post-secondary qualification and 52.85% of participants frequently partook in vigorous activities. According to the Instrumental Activities of Daily Living Indices, 98.09% of participants had relatively low difficulty carrying out instrumental tasks.

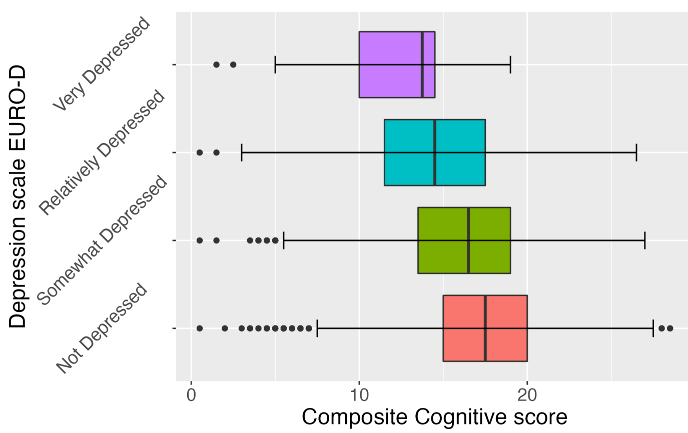
The graphs in fig. 3 show how the cognitive score varies for different levels of some variables from table 2. In general, the boxplots support the background research claims regarding risk factors of dementia.



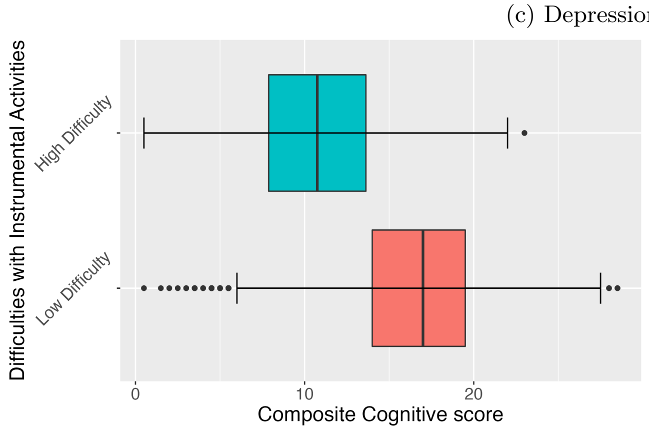

(a) No. of Chronic Diseases



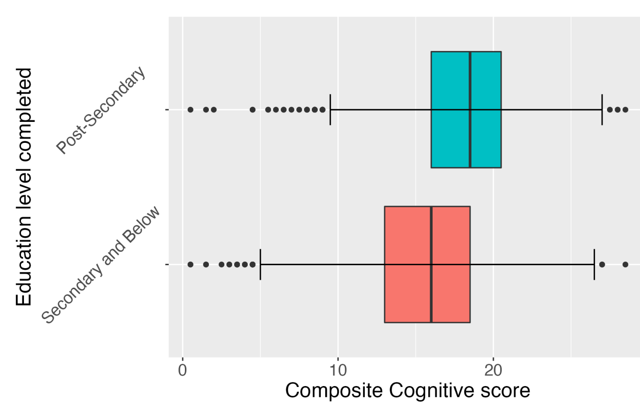

(b) Frequency of Vigorous Activities




(c) Depression Scale EURO-D




(d) Level of Difficulty with Instrumental Activities




(e) Education Level Completed

Figure 3: Graphical representation of the relationships between composite cognitive score and some key variables identified from the model.

In fig. 3a, it was observed that a general trend of an increase in the number of chronic diseases a person had was associated with a decrease in the five-number summary of cognitive score. This was supported by Hassen et al. [6]. However, the minimum, lower quartile and median cognitive score saw an increase for "7-10 chronic diseases". The median and minimum value of cognitive score increased from 15.25 and 3 respectively for "4-6 chronic diseases" to 16.5 and 12 respectively for "7-10 chronic diseases". This deviation from the general trend could be due to a smaller sample size for those with 7-10 chronic diseases (only 0.3% of all observations) compared to the sample sizes of the remaining levels.

Upon analysing fig. 3c, an increase in the severity of depression generally seemed to be associated with a decrease in the five-number summary of cognitive score, with the difference between median cognitive scores of "Not Depressed" and "Very Depressed" being 3.75. The existence of this general trend was also mentioned in Livingston et al. [7]. Also, the lower levels of depression severity seemed to have more outliers, with 57 outliers present for "Not Depressed" and a total of 22 outliers present for the remaining levels combined. While these outliers exist, the proportion of outliers to number of observations was small enough that the outliers may not affect the possible correlation between cognitive score and depression levels.

Considering fig. 3d, those who had relatively lower difficulty performing daily instrumental activities such as making telephone calls or shopping for groceries generally had a higher cognitive score. The difference in median cognitive scores between those experiencing high difficulty and low difficulty was 6.25. There was a noticeable difference in sample sizes between those with high difficulty (96 observations) and low difficulty (4935 observations).

Full five-number summaries for each of the boxplots in fig. 3 are provided in appendix D.

A comparison was also made between sexes to see how an increase in age affected the cognitive score of each sex (fig. 4). It was visible that men generally had a lower cognitive score compared to women for all ages. However, cognitive score decreased at similar rates for increasing age in both sexes.

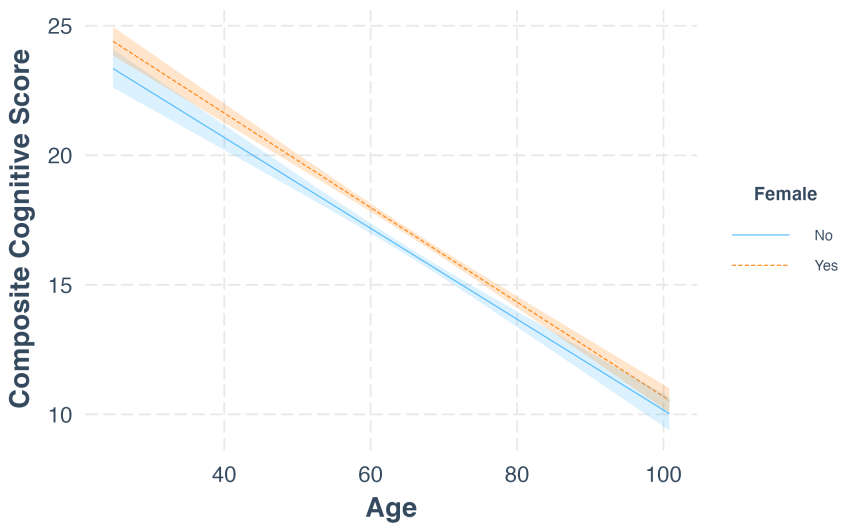



Figure 4: Graph to show how Composite cognitive score is affected with age, and how this compares with the different genders.

## 4.1 Model Results

The key results in table 4 were chosen based on variables which were thought to be the biggest risk factors before constructing the model and also showed a significant relationship with a participant's cognitive score after constructing the model. These included the participant's age and sex, information about where they lived and whether they were born in Estonia. Additionally, variables for information about level of education, ability to complete everyday instrumental activities, how active they are in day-to-day life, and if they received any outside help on a daily basis were also included.

The table includes the name for key explanatory variables $x_i$ or intercept and the coefficient estimate $\hat{\beta}_i$. The $p$-values indicate that the variable was significant in the model. If the $p$-value for $\beta_i$ was lesser than the significance level $\alpha$, we could reject the null-hypothesis $H_0 : \beta_i = 0$. The significance column in table 4 denotes the significance of the variable for the $\alpha$ significance level.

| | Estimate | $p$-value | Significance | Relationship with Composited Cognitive Score |
|---|---|---|---|---|
| Intercept | 23.28 | 0.00 | *** | - |
| female | 1.12 | 0.00 | *** | Positive |
| age | -0.14 | 0.00 | *** | Negative |
| native | 1.07 | 0.00 | *** | Positive |
| loc_Town_rural | -1.37 | 0.00 | *** | Negative |
| edu_Post_Secondary | 1.70 | 0.00 | *** | Positive |
| outside_help | -0.69 | 0.01 | * | Negative |
| iadlza_High_Difficulty | -2.77 | 0.00 | *** | Negative |
| vig_act_Rarely | -0.39 | 0.00 | *** | Negative |
| age:doc_vis_More_than_10 | -0.05 | 0.00 | *** | Negative |

Table 4: Summary statistics for key results of the coefficient parameters of the model along with their relationship with the composite cognitive score. The symbols of significance are defined as *** : $\alpha = 0.001$, ** : $\alpha = 0.01$, and * : $\alpha = 0.05$.

When analysing the coefficients of the explanatory variables in the model, it is worth noting that the variables which had positive (or negative) coefficient estimates had a positive (or negative) association with cognitive score. The results suggested that being female, being born in Estonia, and having a post-secondary education qualification had a positive relationship with cognitive score, where being female increased one's cognitive score by a factor of 1.12 and completing post-secondary education increased the cognitive score by a relatively large factor of 1.7. The remaining variables from table 4 had a negative relationship with cognitive score. To visualise the positive and negative associations of the explanatory variables in table 4, the forest plot in fig. 5 contains coefficient estimates of the variables (denoted by black points) and shows the relationship between variables and cognitive score.

An increase in age by one year reduced the cognitive score only by a factor of 0.14. So, although age had a negative relationship with the cognitive score, the decrease in cognitive score as a person aged was not large compared to other variables, which contradicted previous assumptions that age was the biggest risk factor. However, in checking AICs for the model, it was found that age was still a significant variable compared to others. The largest relative association from the key results was found to be one's inability to perform daily instrumental activities (represented by `iadlza`) with the cognitive score of those who had high difficulty with these tasks decreasing by factor of 2.77. Additionally, living in a rural area, as opposed to living in a city, seemed to decrease the cognitive score by a relatively large factor of 1.37. The intercept estimate was 23.28, implying that if all explanatory variables $x_1 = x_2 = \cdots = x_{18} = 0$ for an observation, the cognitive score would be 23.28.

## 4.2 Model Assessment

The forest plot (fig. 5) displays 95% confidence intervals (CIs), represented by the red horizontal lines, for each of the coefficient estimates. It was observed that the estimates for age and the interaction between age and number of doctor visits had narrow CIs indicating that the uncertainty of the estimate was lower than the estimates for instrumental activities of daily living and whether the participant had received help from outside the household which had wider CIs, meaning there was more uncertainty around the estimate.

Finally, at least one of $\beta_i \neq 0, \quad (i = 0, \ldots, 18)$, since the $p-$value $\approx 0$ for the $F-$statistic, suggesting that null hypothesis $H_0 : \beta_0 = \beta_1 = \cdots = \beta_{18}$ can be rejected.

In a wider context, the above analysis could mean that some variables having a negative association with cognitive score (e.g. age, depression level, etc.) could be potential risk factors. The coefficient estimates, CIs and $p$-values of all explanatory variables are given in table 9 in appendix B.

| Test dataset | Scores of main model | Scores of minimal model |
|---|---|---|
| Test 1: Training data (5031 observations from Estonia in wave 4) | 3.43 | 3.82 |
| Test 2: 200 observations from Estonia in wave 4 | 3.52 | 4.00 |
| Test 3: Data from Austria in wave 4 | 3.77 | 4.17 |
| Test 4: Data from Estonia in wave 6 | 3.90 | 4.21 |

Table 5: Table of RMSE scores for the main model and minimal model on the four test datasets.

To assess the predictive performance of the model, RMSE scores were calculated to compare the accuracy of predicted cognitive scores for different datasets. The model was built using the training dataset then tested on the four test datasets described in subsection 3.3. A lower RMSE score indicated a higher predictive performance of the model on the dataset. From table 5, it was observed that for the main model alone, the RMSE scores for the training dataset were highest, which was expected since the model was built using this dataset. The predictive performance decreased when the model was applied to the Test 2 dataset. A further increase in RMSE scores for the third and fourth test dataset was observed. When considering the minimal model alone, the same trend of a decrease in predictive performance due to an increase in RMSE scores was observed across the four datasets. However, a key result was that the predictive performance of the main model was higher than the minimal model where cognitive score depends on age alone, since the minimal model has higher RMSE scores than the main model for each dataset. This suggested that the explanatory variables do have some importance in predicting cognitive scores. The increase in RMSE scores across the datasets was not too large, so the main model may be applied to different datasets.

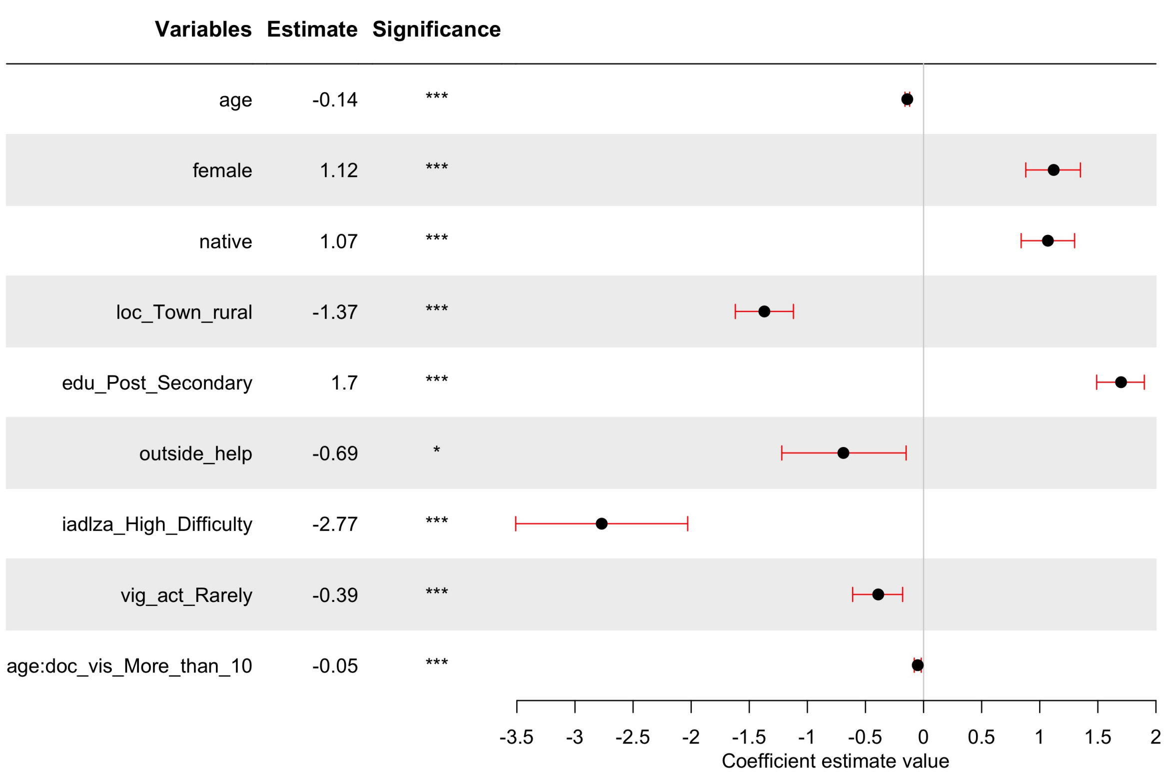



Figure 5: Forest plot to visualise the coefficients for key explanatory variables and their corresponding confidence intervals.

# 5 Conclusion

The analysis of the observational data could suggest that some of the risk factors of dementia include age, difficulty in instrumental activities of daily living, lack of vigorous activities and higher depression levels, amongst others, since they could decrease an individual's cognitive score. This was somewhat inline with the background research suggesting that dementia is associated with an increase in age, with depression also being a risk factor. It was seen that a lack of vigorous activity and higher difficulty in instrumental activity could be possible risk factors since they caused a decrease in cognitive score by a factor of 0.39 and 2.77 respectively. It was also observed that females generally had higher cognitive scores than males in a given age group, though it could not be concluded that males were at a higher risk of dementia than females.

While the analysis did support some of our assumptions from the background research on risk factors of dementia, there were some risk factors that we were not able to explore due to a lack of information regarding these risk factors (e.g., smoking and drinking habits, childhood health conditions, mathematical abilities at age 10, etc.). Also, this analysis was carried out using observational data from one country (Estonia) in a single wave of interviews (wave 4) as this has a larger sample size than some of the other countries. A lack of consistency in the questions asked during the data collection period and large amounts of missing data made the data difficult to extrapolate to varied populations. So, while the statistical model has been assessed to investigate possible risk factors of dementia for different datasets, there may be some factors that have not been investigated as intended.

## References


[1] Axel Börsch-Supan. *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 4*. en. 2022. DOI: 10.6103/SHARE.W4.800. URL: http://www.share-project.org/data-documentation/waves-overview/wave-4.html.

[2] Axel Börsch-Supan. *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 6*. en. 2022. DOI: 10.6103/SHARE.W6.800. URL: http://www.share-project.org/data-documentation/waves-overview/wave-6.html.

[3] Axel Börsch-Supan and Stefan Gruber. *easySHARE*. 2022. DOI: 10.6103/SHARE.EASY.800. URL: http://www.share-project.org/special-data-sets/easyshare.html.

[4] Mayo Clinic. *Dementia*. URL: https://www.mayoclinic.org/diseases-conditions/dementia/symptoms-causes/syc-20352013.

[5] Centers for Disease Control and Prevention. *What is Dementia?* Apr. 2019. URL: https://www.cdc.gov/aging/dementia/index.html#:~:text=Dementia%5C%20is%5C%20not%5C%20a%5C%20specific.

[6] Céline Ben Hassen et al. "Association between age at onset of multimorbidity and incidence of dementia: 30 year follow-up in Whitehall II prospective cohort study". In: *BMJ* 376 (Feb. 2022), e068005. DOI: 10.1136/bmj-2021-068005. URL: https://www.bmj.com/content/376/bmj-2021-068005.

[7] Gill Livingston et al. "Dementia prevention, intervention, and care: 2020 report of the Lancet Commission". In: *The Lancet* 396.10248 (2020), pp. 413–446. ISSN: 0140-6736. DOI: https://doi.org/10.1016/S0140-6736(20)30367-6. URL: https://www.sciencedirect.com/science/article/pii/S0140673620303676.

[8] NHS. *Dementia guide*. Dec. 2017. URL: https://www.nhs.uk/conditions/dementia/.

[9] Alzheimer's Society. *What Is dementia? — Alzheimer's Society*. 2021. URL: https://www.alzheimers.org.uk/about-dementia/types-dementia/what-is-dementia.

# 6 Appendix

## A Summary Statistics for training data (5031 observation from wave 4 in Estonia)

| Cognitive score | Age | Job situation | Number of chronic diseases | Education level completed | Quality of life (CASP-12 score) |
|---|---|---|---|---|---|
| Min. : 0.50 | Min. : 24.80 | Retired :2780 | 0 :1365 | Secondary and Below:3162 | Low : 275 |
| 1st Qu.:14.00 | 1st Qu.: 57.80 | Employed :1697 | 1-3 :3183 | Post-Secondary :1869 | Moderate:2412 |
| Median :17.00 | Median : 65.50 | Unemployed : 223 | 4-6 : 468 | Other : 0 | High :2344 |
| Mean :16.58 | Mean : 66.03 | Sick_Disabled: 277 | 7-10: 15 | | |
| 3rd Qu.:19.50 | 3rd Qu.: 73.70 | Homemaker : 47 | | | |
| Max. :28.50 | Max. :100.80 | Other : 7 | | | |

Table 6: Table containing summary statistics for continuous variables cognitive score and age as well as sample size per factor level for categorical variables - job situation, number of chronic diseases, education level completed, and the quality of life (CASP-12 score) - from the data from Estonia in wave 4.

| Female | Born in country of survey | Outside help received | Stayed in a nursing home in the past year | Location of survey | Maximum grip strength |
|---|---|---|---|---|---|
| No :1986 | No :1257 | No :3964 | No :5020 | City_sub :1196 | Low :2106 |
| Yes:3045 | Yes:3774 | Yes:1067 | Yes: 11 | Town_rural:3835 | High:2925 |

Table 7: Table containing summary statistics for sample size per level for categorical variables - female, born in the country of survey, outside help received, stayed in a nursing home in the past year, location of survey, and maximum grip strength - from the data from Estonia in wave 4.

| Instrumental activities | Number of doctor visits | Frequency of vigorous activities |
|---|---|---|
| Low Difficulty :4935 | 0-10 :4389 | Often :2659 |
| High Difficulty: 96 | More than 10: 642 | Rarely:2372 |

Table 8: Table containing summary statistics for sample size per level for categorical variables - instrumental activities, number of doctor visits and the frequency of vigorous activities - from the data from Estonia in wave 4.

# B Summary Statistics Table For All Variables In The Model

| | Estimate | CI_lower | CI_upper | p.value | Significance |
|---|---|---|---|---|---|
| Intercept | 23.28 | 22.01 | 24.55 | 0.00 | *** |
| age | -0.14 | -0.16 | -0.12 | 0.00 | *** |
| female | 1.12 | 0.88 | 1.35 | 0.00 | *** |
| native | 1.07 | 0.84 | 1.30 | 0.00 | *** |
| loc_Town_rural | -1.37 | -1.62 | -1.12 | 0.00 | *** |
| edu_Post_Secondary | 1.70 | 1.49 | 1.90 | 0.00 | *** |
| outside_help | -0.69 | -1.22 | -0.15 | 0.01 | * |
| chronic_mod1-3 | 0.26 | 0.03 | 0.49 | 0.03 | * |
| chronic_mod4-6 | 0.36 | -0.03 | 0.76 | 0.07 | . |
| chronic_mod7-10 | 2.99 | 1.22 | 4.77 | 0.00 | *** |
| life_quality_Moderate | 1.37 | 0.91 | 1.83 | 0.00 | *** |
| life_quality_High | 1.68 | 1.19 | 2.16 | 0.00 | *** |
| eurod_Somewhat_Depressed | -0.52 | -0.75 | -0.29 | 0.00 | *** |
| eurod_Relatively_Depressed | -1.05 | -1.45 | -0.65 | 0.00 | *** |
| eurod_Very_Depressed | -2.36 | -3.60 | -1.12 | 0.00 | *** |
| doc_vis_More_than_10 | 3.75 | 1.69 | 5.82 | 0.00 | *** |
| nursing | -2.35 | -5.11 | 0.42 | 0.10 | . |
| iadlza_High_Difficulty | -2.77 | -3.51 | -2.03 | 0.00 | *** |
| maxgrip_High | 0.61 | 0.35 | 0.86 | 0.00 | *** |
| vig_act_Rarely | -0.39 | -0.61 | -0.18 | 0.00 | *** |
| job_situ_Employed | 0.27 | -0.03 | 0.57 | 0.08 | . |
| job_situ_Unemployed | -0.40 | -0.93 | 0.14 | 0.15 | |
| job_situ_Sick_Disabled | -1.17 | -1.66 | -0.68 | 0.00 | *** |
| job_situ_Homemaker | -0.56 | -1.57 | 0.46 | 0.28 | |
| job_situ_Other | -0.38 | -2.94 | 2.19 | 0.77 | |
| loc_Town_rural:outside_help | 0.68 | 0.09 | 1.27 | 0.02 | * |
| age:doc_vis_More_than_10 | -0.05 | -0.08 | -0.02 | 0.00 | *** |
| nursing:iadlza_High_Difficulty | 7.24 | 3.08 | 11.40 | 0.00 | *** |

Table 9: All the summary statistics for the different variables and their levels considered in the model. The symbols of significance are defined as *** : $\alpha = 0.001$, ** : $\alpha = 0.01$, * : $\alpha = 0.05$, and . : $\alpha = 0.1$, with no symbol showing no significance for $\alpha \leq 0.1$.

# C Plots for Checking Model Assumptions

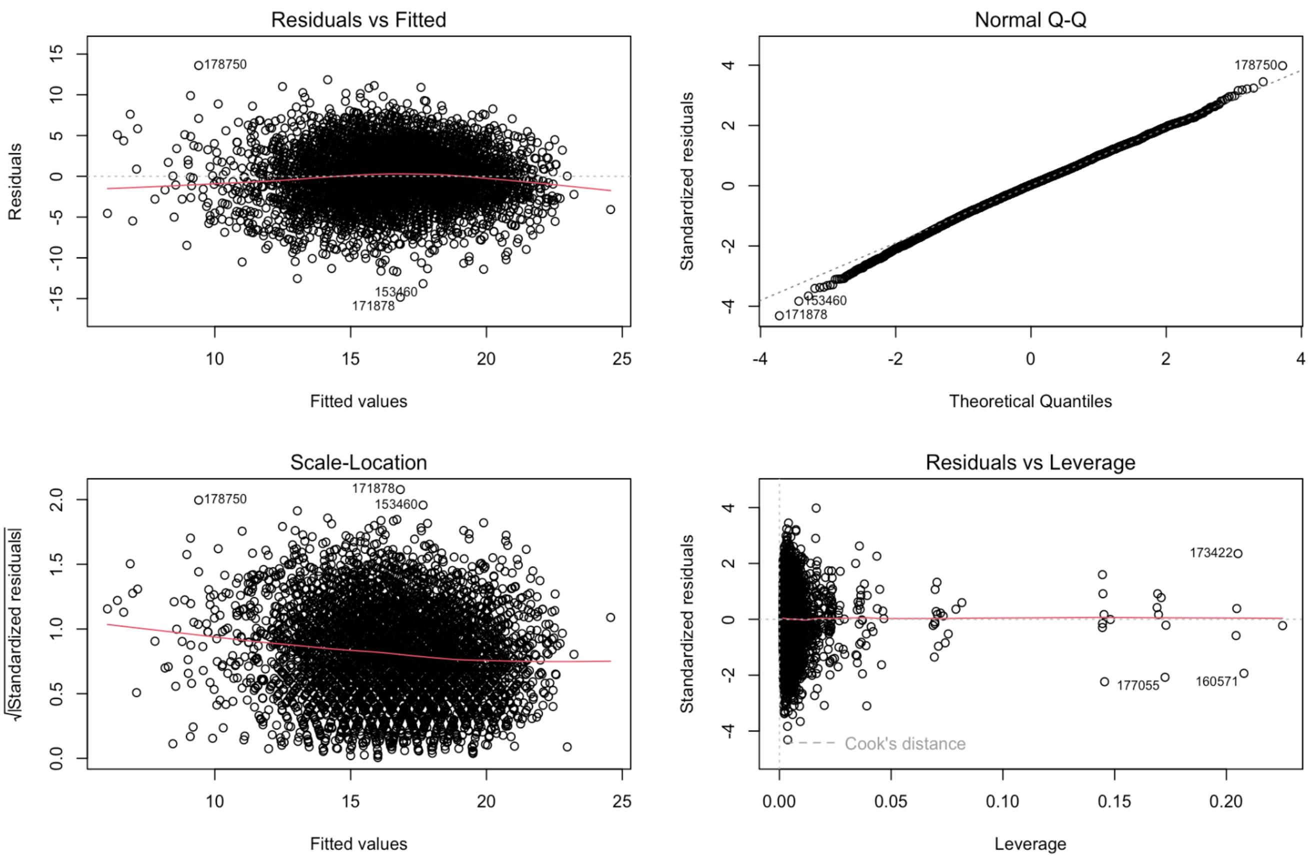



Figure 6: The plots for regression diagnostic on the final model.

# D Tables of Boxplot Summary Statistics for fig. 3

| Level | Minimum | Lower Quartile | Median | Upper Quartile | Maximum | Number of Observations | Number of Outliers |
|---|---|---|---|---|---|---|---|
| 0 | 7.50 | 15.00 | 18.00 | 20.00 | 27.00 | 1365 | 27 |
| 1-3 | 6.00 | 14.00 | 16.50 | 19.50 | 27.50 | 3138 | 43 |
| 4-6 | 3.00 | 12.00 | 15.25 | 18.12 | 26.50 | 468 | 2 |
| 7-10 | 12.00 | 14.75 | 16.50 | 17.75 | 21.50 | 15 | 1 |

Table 10: Table of boxplot summary statistics for composite cognitive score with the number of chronic diseases, fig. 3a.

| Level | Minimum | Lower Quartile | Median | Upper Quartile | Maximum | Number of Observations | Number of Outliers |
|---|---|---|---|---|---|---|---|
| Often | 7.50 | 15.00 | 17.50 | 20.00 | 27.50 | 2659 | 27 |
| Rarely | 4.00 | 13.00 | 16.00 | 19.00 | 27.00 | 2372 | 15 |

Table 11: Table of boxplot summary statistics for composite cognitive score with the frequency of vigorous activities, fig. 3b.

| Level | Minimum | Lower Quartile | Median | Upper Quartile | Maximum | Number of Observations | Number of Outliers |
|---|---|---|---|---|---|---|---|
| Not Depressed | 7.50 | 15.00 | 17.50 | 20.00 | 27.50 | 3038 | 57 |
| Somewhat Depressed | 5.50 | 13.50 | 16.50 | 19.00 | 27.00 | 1566 | 18 |
| Relatively Depressed | 3.00 | 11.50 | 14.50 | 17.50 | 26.50 | 397 | 2 |
| Very Depressed | 5.00 | 10.00 | 13.75 | 14.50 | 19.00 | 30 | 2 |

Table 12: Table of boxplot summary statistics for composite cognitive score with the depression scale EURO-D, fig. 3c.

| Level | Minimum | Lower Quartile | Median | Upper Quartile | Maximum | Number of Observations | Number of Outliers |
|---|---|---|---|---|---|---|---|
| Low Difficulty | 6.00 | 14.00 | 17.00 | 19.50 | 27.50 | 4935 | 61 |
| High Difficulty | 0.50 | 7.88 | 10.75 | 13.62 | 22.00 | 96 | 1 |

Table 13: Table of boxplot summary statistics for composite cognitive score with the difficulties participants have with instrumental activities, fig. 3d.

| Level | Minimum | Lower Quartile | Median | Upper Quartile | Maximum | Number of Observations | Number of Outliers |
|---|---|---|---|---|---|---|---|
| Secondary and Below | 5.00 | 13.00 | 16.00 | 18.50 | 26.50 | 3162 | 32 |
| Post-Secondary | 9.50 | 16.00 | 18.50 | 20.50 | 27.00 | 1869 | 39 |

Table 14: Table of boxplot summary statistics for composite cognitive score with the level of education completed, fig. 3e.